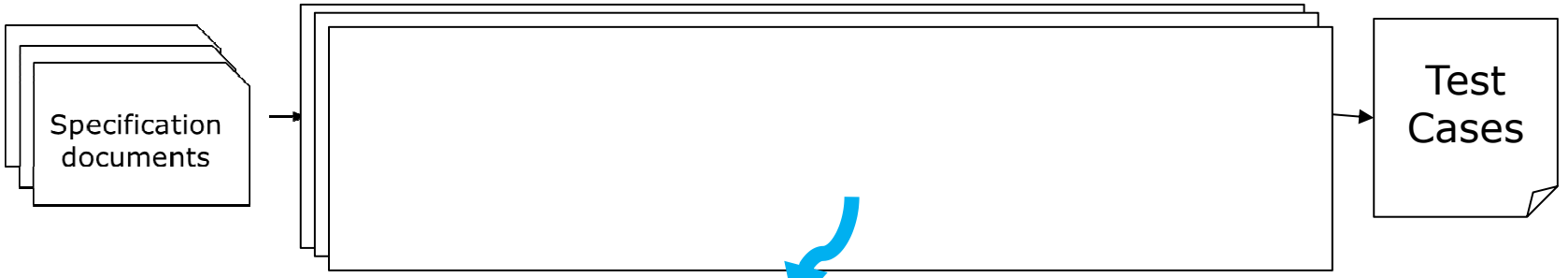
IBM

# Semantic Analysis Technique of Logics Retrieval for Software Testing from Specification Documents

*Satoshi Masuda, Futoshi Iwama, Nobuhiro Hosokawa*, IBM Research – Tokyo
*Tohru Matsuodani*, Debug Engineering Research Laboratory
*Kazuhiko Tsuda*, University of Tsukuba

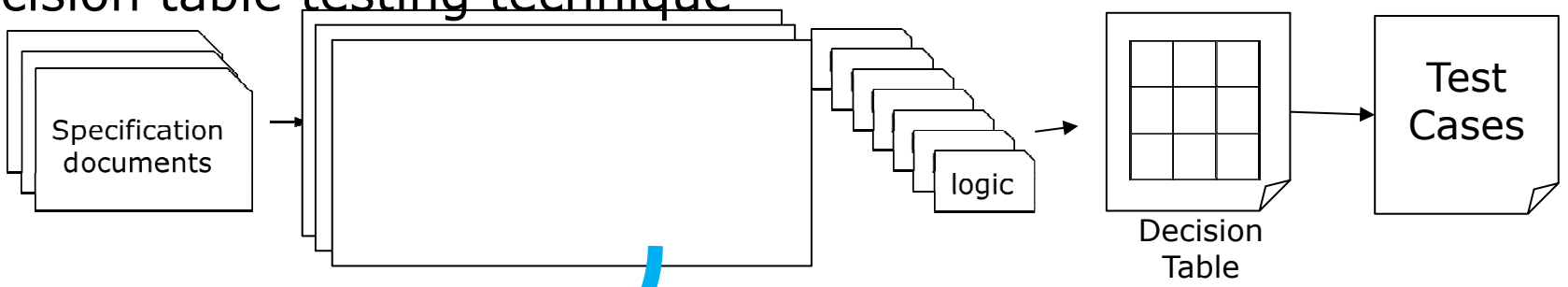
# I. Introduction

-Create test cases from specification documents



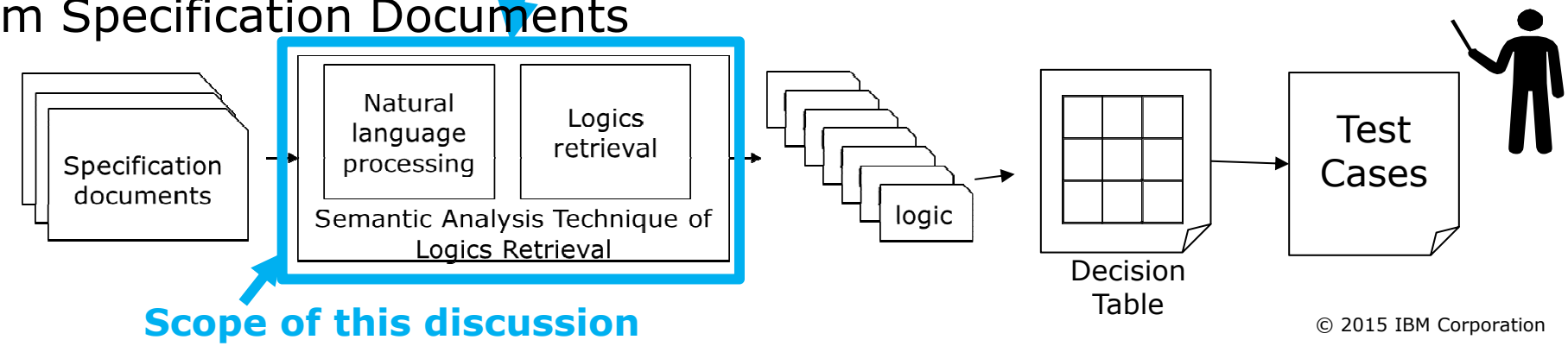

One of techniques

-Decision table testing technique



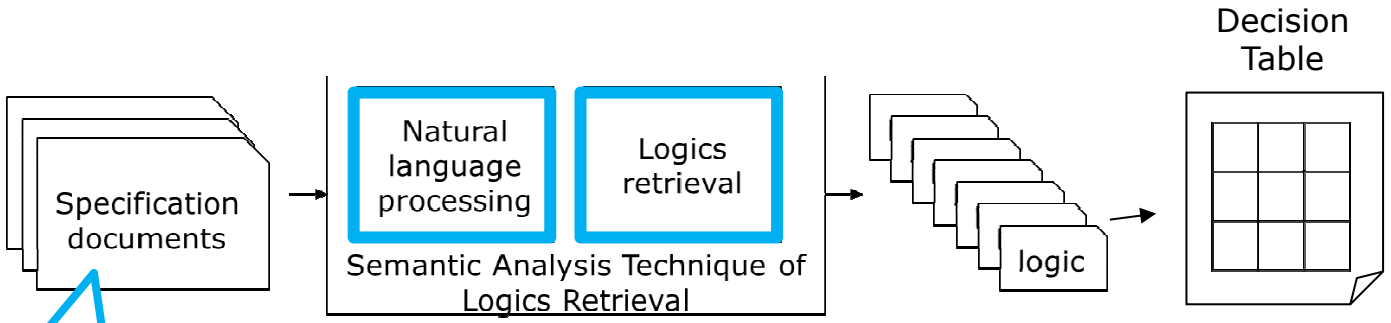

One of techniques

-Semantic Analysis Technique of Logics Retrieval for Software Testing from Specification Documents




Scope of this discussion

# II. Background and Motivation

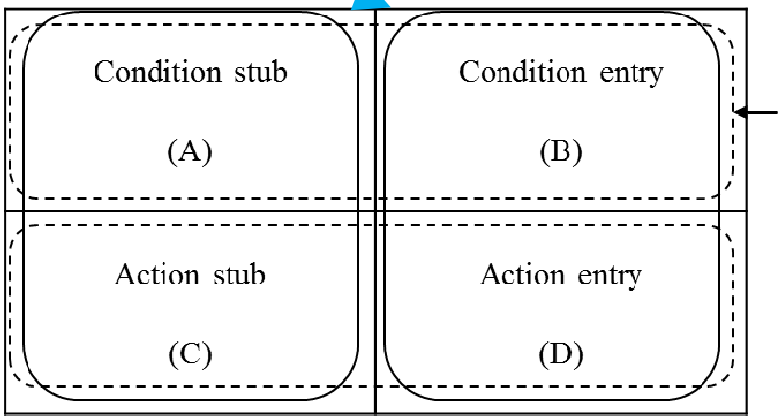



**Stakeholders** often **use their natural language** to exchange their idea, business processes, business rules and other specifications and **describe the specifications into documents**.

logic

*if (A) is (B), (C) is (D).*

Decision Table



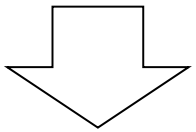

**How to retrieve logics from specification documents?**

**Japanese 日本語**

# III. Related works

- Natural Language Processing
  - Many works have been already done, even in Japanese Language.
    - Morphological analysis
    - Structural analysis (Dependency analysis)
    - Semantic analysis

    etc..

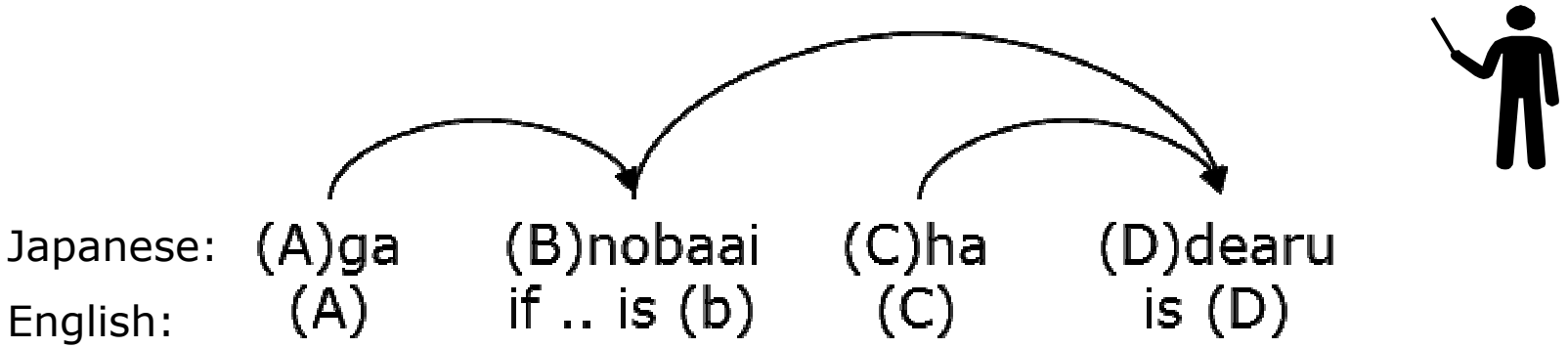


We use the results of NLP techniques.

# III. Related works

- **Sneed2007** [2] presented testing against natural language requirements. The approach was to analyze requirements and extract test cases from them.
  - The paper did not mention about techniques of natural language processing and the target language is English.
- **Saeki1989** [11] presented software development process from natural language specification. That was an approach to solve problems about natural language specification by the process which was defined as “design” and “elaborate”.
  - The approach was not to try to automate by using techniques natural language processing.

- Logic retrieval




- Sample specification sentence

1. Japanese:

   *"Miraini betsuno detaga haitteitabaai, sono jitennno tyokuzennwo shuuryoubitosurukoto."*

2. English:

   *"If another data exists in a future field, set a date just before the data as end date."*

# Results of dependency analysis for the sample sentence

| $i$ | $P_m(i)$ | $Dep_m(i)$ |
|---|---|---|
| 1 | miraini (future field) | 4 |
| 2 | betsuno (another) | 3 |
| 3 | detaga (data) | 4 |
| 4 | haitteita**baai** (**if** exist) | 8 |
| 5 | sono (the) | 6 |
| 6 | jitenno (a date) | 7 |
| 7 | tyokuzennwo (just before) | 8 |
| 8 | shuuryoubitosurukoto (sets end date) | ***T*** |

# IV. SEMANTIC ANALYSIS TECHNIQUE OF LOGICS RETRIEVAL

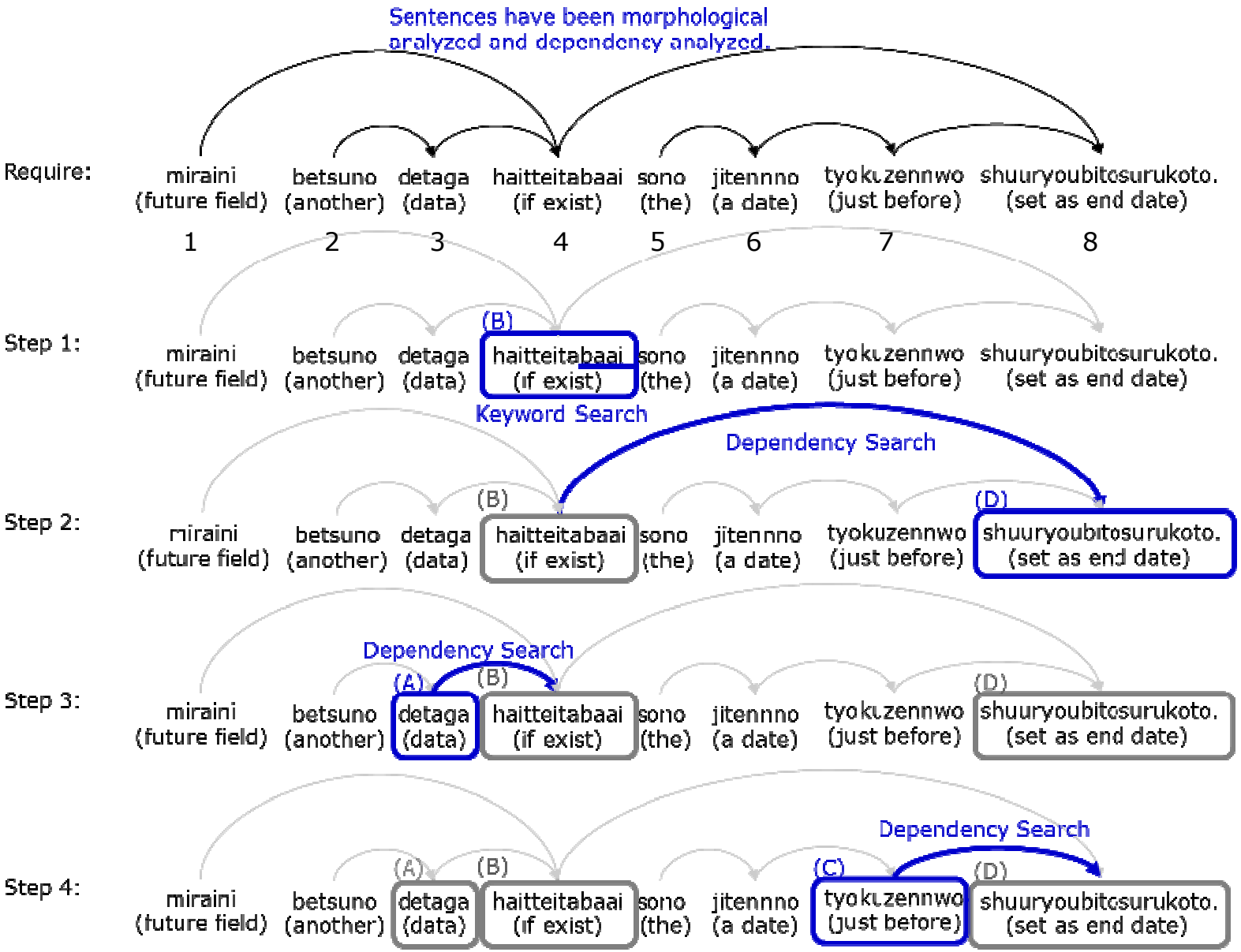



Fig. 3. Steps to retrieve logics from sentences.

- Require: Sentences have been morphological analyzed and dependency parsed as Table I.
- Step 1: Search (B) words by keyword pattern matching. In this paper we set a rule that (B) Words have “baai” in Japanese which means “if” in English.
- Step 2: Search (D) words which (B) depend on.
- Step 3: Search nearest (C) word which depends on (D)
  - Nearest word has strongest dependency with the target word.
- Step 4: Search nearest (A) word which depend on (B)
  - As same as step 3, nearest word has strongest dependency with the target word.

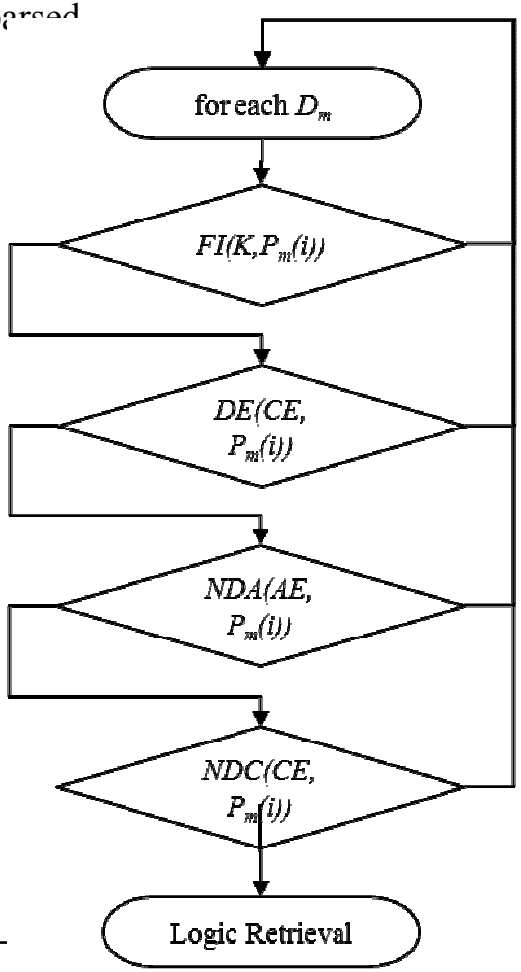
# V. Experiments

We implemented the algorithm on Perl programs and made experiments

**Algorithm 1** The analysis technique algorithm

**Input:** documents which have been morphological analyzed and dependency parsed

```
1:  for each D_m do
2:    for all P_m(i) do
3:      if P_m(i) = K1 then
4:         CEi = i
5:        P_m(i) = CE
6:        if Pm(Dep_m(i)) ∈ D_m and P_m(Dep_m(i)) ∈ T then
7:           AEi = i
8:          P_m(Dep_m(i)) = AE
9:            else next D_m
10:         end next D_m
11:   end for
12:   for all P_m(i) do
13:     if Dep_m(i) = CEi and max (i) then
14:        P_m(Dep_m(i)) = AS
15:     else next D_m
16:     if Dep_m(i) = AEi and max(i) then
17:       P_m(Dep_m(i)) = CS
18:     else next D_m
19:   end for
20: end for
```



# V. Experiments

## Specification documents in Japanese

| Doc Groups | Pages | Characters (Double Bytes) | File Size (Bytes) | Create date |
|---|---|---|---|---|
| A | 93 | 74,559 | 1,702,400 | 2010/09/16 |
| A | 17 | 3,283 | 4,555,776 | 2014/03/18 |
| A | 14 | 9,502 | 60,416 | 2008/06/01 |
| A | 15 | 11,182 | 51,712 | 2014/07/01 |
| A | 76 | 63,908 | 1,008,129 | 2003/03/01 |
| B | 10 | 6,825 | 50,688 | 2014/06/01 |
| B | 2 | 779 | 97,280 | 2007/11/14 |
| B | 22 | 14,560 | 499,712 | 2008/04/01 |
| B | 17 | 11,211 | 54,272 | 2006/09/01 |
| B | 19 | 6,844 | 79,360 | 2010/05/01 |
| B | 9 | 7,844 | 140,800 | 2014/01/01 |
| C | 157 | 152,568 | 1,995,521 | 2012/11/01 |
| C | 56 | 55,738 | 609,407 | 2005/07/01 |
| C | 73 | 70,167 | 2,775,933 | 2011/07/06 |
| D | 11 | 9,060 | 37,888 | 2012/04/01 |
| D | 41 | 38,527 | 653,312 | 2012/04/01 |
| D | 327 | 296,451 | 6,210,032 | 2012/04/01 |
| E | 25 | 20,364 | 880,640 | 2009/12/02 |
| E | 8 | 6,717 | 77,312 | 2012/04/01 |
| E | 7 | 9,852 | 313,871 | 2014/04/01 |
| E | 10 | 8,035 | 72,192 | 2013/05/01 |
| F | 10 | 4,419 | 90,624 | 2011/10/01 |
| F | 8 | 5,012 | 10,223 | 2013/04/01 |
| F | 173 | 188,919 | 1,172,389 | 2007/04/01 |
| F | 18 | 26,466 | 618,214 | 2013/04/01 |

# V. Experiments

TABLE IV. RESULTS THE ANALYSIS TECHNIQUE VS. EVALUATION

| **The analysis technique** | ***Positive*** | | ***Negative*** | |
|---|---|---|---|---|
| **Evaluations** | ***Positive (a)*** | ***Negative (b)*** | ***Positive (c)*** | ***Negative (d)*** |
| A | 31 | 1 | 15 | 2 |
| B | 15 | 1 | 4 | 3 |
| C | 43 | 2 | 17 | 4 |
| D | 62 | 5 | 33 | 21 |
| E | 35 | 1 | 19 | 6 |
| F | 107 | 8 | 40 | 26 |

TABLE V. RESULTS OF RECALL AND PRECISION

| | **Document Groups** | | | | | |
|---|---|---|---|---|---|---|
| | ***A*** | ***B*** | ***C*** | ***D*** | ***E*** | ***F*** |
| Precision | 0.97 | 0.94 | 0.96 | 0.93 | 0.97 | 0.93 |
| Recall | 0.67 | 0.79 | 0.72 | 0.65 | 0.65 | 0.73 |

# VI. Conclusion

- We proposed the analysis technique, a semantic analysis technique of logics retrieval for software testing from Japanese public sector's specification documents.
- The result was that the precision reached 0.93 to 0.97 and recall reached 0.65 to 0.79.
- We confirmed the analysis technique could retrieve logics from Japanese natural language specification documents.
- This result is the starting point to research about harmonization between natural language processing and software testing.

# References


1. Japan Industry Standard Association, "JIS X 0125-1986 Decision Table", in Japan Industry Standards, 1986
2. H. M. Sneed, "Testing against natural language requirements", in IEEEQuality Software, 2007. QSIC'07. Seventh International Conference on, 2007, pp.380-387
3. K. Tsuda, and M. Nakamura, "The extraction method of the word meaning class", in IEEE Knowledge-Based Intelligent Information Engineering Systems, 1999. Third International Conference, 1999, pp.534-537
4. T. Matsodani, "Application of Decision Table to Software Logic with Designing and Testing", in Reliability Enginerring Association of Japan, 2012, pp.397-404
5. K. Uetsuki, T. Matsuodani, and K. Tsuda, "An efficient software testing method by decision table verification", in International Journal of Computer Applications in Technology, 2013, pp.54-64
6. S. Kishimoto, "Government procurement system and IT system -Who the raised or IT general constructor", in http://www.rieti.go.jp/jp/projects/denshi/04.pdf, 2003, pp.37-47
7. I. C. S. S. E. S. Committee, and I.-S. S. Board, "IEEE Recommended Practice for Software Requirements Specifications", in Institute of Electrical and Electronics Engineers, 1998
8. N. Fukui, "Deriving the differences between English and Japanese ", in English Linguistics, 1988, pp.249-270
9. C. Kim, S.-M. Kim, and K.-W. Song, "Measurement of Level of Quality Control Activities in Software Development [Quality Control Scorecards]", in IEEEConvergence and Hybrid Information Technology, 2008. ICHIT'08. International Conference on, 2008, pp.763-770
10. S. KUNO, "The structure of the Japanese language", in Cambridge, MA:MIT Press., 1973
11. M. Saeki, H. Horai, and H. Enomoto, "Software development process from natural language specification", in ACMProceedings of the 11th international conference on Software engineering, 1989, pp.64-73
12. A. Shhimazu, "Legal Engineering: Methodology for Designing Trustworthy Social Systems -Legal Document Analysis-", in IEICE Foundamentals Review, 2012, pp.320-328
13. T. Kudo and Y. Matsumoto, "Japanese Dependency Analysis Using Cascaded Chunking", in Journal of Information Processing Society of Japan, 2002, pp.1834-1842
14. Y. Kimura, M. Nakamura, A. Shimazu, "Trasnformation logical expression from legal documents which have listing and references", in Assocaiton for natural Language Processing, 2008, pp.612-615
15. Y. Tang, and R. Meersman,"On constructing semantic decision tables", in SpringerDatabase and Expert Systems Applications,2007,pp.34-44
16. K. Tanaka, I. Kawazoe and H. Narita, "Standard structure of legal provisions - For the leagal knowledge processing by natural language -", in Information Processing Society of Japan Natural Language Processing, 1993, pp.79-86
17. T. Katayama, "Legal engineering-an engineering approach to laws in e-society age", in Proc. of the 1st Intl. Workshop on JURISIN, 2007
18. S. Masuda, T. Matsuodani, and K. Tsuda,"A Method of Creating Testing Pattern for Pair-wise Method by Using Knowledge of Parameter Values", in Procedia Computer Science,2013,pp.521-528

# The end of presentation